

RETINAL-INSPIRED FILTERING FOR DYNAMIC IMAGE CODING

Effrosyni Doutsi^{1,2}, Lionel Fillatre¹, Marc Antonini¹ and Julien Gaulmin²

¹Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis, France

²4G-TECHNOLOGY, 460 avenue de la Quiera 06370 Mouans Sartoux - France.

ABSTRACT

This paper introduces a novel non-Separable sPAtioteMporal filter (non-SPAM) which enables the spatiotemporal decomposition of a still-image. The construction of this filter is inspired by the model of the retina which is able to selectively transmit information to the brain. The non-SPAM filter mimics the retinal-way to extract necessary information for a dynamic encoding/decoding system. We applied the non-SPAM filter on a still image which is flashed for a long time. We prove that the non-SPAM filter decomposes the still image over a set of time-varying difference of Gaussians, which form a frame. We simulate the analysis and synthesis system based on this frame. This system results in a progressive reconstruction of the input image. Both the theoretical and numerical results show that the quality of the reconstruction improves while the time increases.

Index Terms— Bio-inspired processing, dynamic encoder, non-separable spatiotemporal filter, frame, dual frame.

1. INTRODUCTION

During the last decades there has been a great progress in image and video compression standards which enables to deal with the current needs of coding and decoding High Definition (HD) and Ultra High Definition (UHD) signals [1]. However, this progression rate seems to be very low comparing to the increasing rate of the amount of data which needs to be transmitted or stored. For instance, H.265/HEVC [2] was released in 2013, ten years later than the previous standard H.264/AVC [3, 4]. This lack of synchronization restricts their evolution which is currently based on parameterizations and/or improvements of the basic architectures instead of the proposition of groundbreaking approaches.

In this paper we study an alternative compression and decompression model, which is based on the behavior of the visual system. The way the luminance of light is captured, transformed and compressed by the inner part of the eye, the retina, seems to follow the basic principles of compression. The retinal function has been explicitly modeled by neuro-scientists and the experimental results have shown that this should be a very efficient “compression” model [5]. This is

due to the fact that the retina is a layered structure of different kinds of cells. The amount of cells decreases while they are closer to the optic nerve. This structure succeeds in encoding the information in order to fit the limited capacity of the optic nerve. The signal which reaches the eyes is successfully transmitted to the brain despite this bottleneck [6, 7].

Our goal is to study the retinal-inspired transformation from the signal processing point of view and to set the basis for our future bio-inspired dynamic codec. The first attempt in modeling this kind of filter was proposed in [8]. This is a separable spatiotemporal filter structured as a Difference of Gaussians (DoG) pyramid based on [9, 10]. Each layer of this structure is delayed with an exponential temporal function. We improve the filter by explicitly taking into account the time in the design of our non-Separable sPAtioteMporal (non-SPAM) filter.

In coding/decoding systems, the transformation of the signal is extremely important because it results in a more suitable representation in terms of the number of informative coefficients. There are different kinds of transformations like Discrete Fourier Transform (DFT) [11], Discrete Cosine Transform (DCT) [12] and the Discrete Wavelet Transform (DWT) [13, 14, 15, 16], which are currently used in most of the lossy conventional compression standards (i.e., JPEG and JPEG2000). In addition, there are other kinds of filters which have been built in order to serve not only image compression but the general aspects of image and video processing. For instance, the Gaussian and Laplacian pyramids [9] are scaled spatial filters which allow a progressive transmission of a signal. Other kinds of pyramids are for instance the oriented pyramids using Gabor functions [17] or the orthogonal pyramids [18]. Many of these approaches taking into account the time has been extended into spatio-temporal filters which have been applied especially in video surveillance and object tracking techniques [19, 20].

In section 2 we aim to introduce the non-SPAM filter and explain its bio-inspired nature. We also applied this filter on a still image which is flashed for a long time. Then, we prove that the non-SPAM filter has a frame structure in section 3. In section 4, we propose a progressive reconstruction of the input signal which is numerically illustrated in section 5. In the last section, we conclude the paper with a discussion about the future work.

2. NON-SPAM FILTER

The aim of this section is to introduce the non-SPAM filter and to study its behavior. This filter is inspired by the mechanism of photo-receptors and horizontal cells which lie inside the retina (the innest part of the eye). These cells act as edge detectors and at the same time as motion detectors due to the way they connect to each other. These are the features that the non-SPAM filter tries to mimic having a spatial behavior which varies with respect to time. The non-separability of space and time enables the filter to detect temporal variations of luminance even in a uniform spatial region. This is not the case for a separable spatiotemporal filter. We study the non-SPAM filter for the analysis and synthesis of an image $f(x, t)$, where $x \in \mathbb{R}^2$ and $t \in \mathbb{R}^+$, which is observed during a certain time interval. The spatiotemporal convolution of the non-SPAM and the input image results in the function $A(x, t)$ called the activation degree:

$$A(x, t) = K(x, t) \stackrel{x,t}{*} f(x, t) \quad (1)$$

where $\stackrel{x,t}{*}$ is the convolution with respect to space and time. We are going to define now the non-SPAM filter in continuous time and space as following:

$$K(x, t) = C(x, t) - S(x, t), \quad (2)$$

where $C(x, t)$ and $S(x, t)$ are the center and the surround spatiotemporal filters given by equations (3) and (4) respectively:

$$C(x, t) = w_c G_{\sigma_c}(x) T(t), \quad (3)$$

$$S(x, t) = w_s G_{\sigma_s}(x) \left(T * E_{\tau_S} \right)(t), \quad (4)$$

where w_c and w_s are constant parameters, G_{σ_c} and G_{σ_s} are spatial Gaussian filters standing for the center and surround areas respectively, and E_{τ_S} is an exponential temporal filter. The center temporal filter $T(t)$ is given by:

$$T(t) = E_{\tau_{G,n}} \stackrel{t}{*} (\delta_0 - w_c E_{\tau_C})(t), \quad (5)$$

where the gamma temporal filter $E_{\tau_{G,n}}(t)$ is defined by

$$E_{\tau,n}(t) = \frac{t^n \exp(-t/\tau)}{\tau^{n+1}}, \quad (6)$$

where $n \in \mathbb{N}$ and τ is a constant parameter ($E_{\tau,n}(t) = 0$ for $t < 0$), E_{τ_C} is an exponential temporal filter, δ_0 is the dirac function and $\stackrel{t}{*}$ stands for the temporal convolution. In case that $n = 0$, the gamma filter turns to an exponential temporal filter. The convolution of the temporal filter $T(t)$ with the exponential filter E_{τ_S} is related to the delay in the appearance of the surround temporal filter with respect to the center one.

The input signal is a still-image which exists for a long time, hence $f(x, t) = f(x) \mathbb{1}_{[0,\infty]}(t)$ where $f(x)$ is the still-image and $\mathbb{1}$ is the indicator function such that $\mathbb{1}_{[0,\infty]}(t) = 1$, if $0 \leq t \leq \infty$, otherwise 0. In this case, we obtain the following simplified convolution formula.

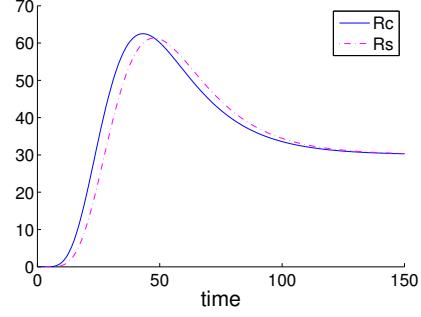


Fig. 1: Temporal filters $R_c(t)$ and $R_s(t)$.

Proposition 2.1. *For a still-image $f(x, t) = f(x) \mathbb{1}_{[0,\infty]}(t)$, the spatiotemporal convolution (1) turns into a spatial convolution:*

$$A(x, t) = \phi(x, t) \stackrel{x}{*} f(x), \quad (7)$$

where $\phi(x, t)$ is a spatial DoG filter weighted by two temporal filters $R_c(t)$ and $R_s(t)$:

$$\phi(x, t) = w_c R_c(t) G_{\sigma_c}(x) - w_s R_s(t) G_{\sigma_s}(x), \quad (8)$$

$$R_c(t) = \int_{t'=0}^t T(t-t') dt', \quad (9)$$

$$R_s(t) = \int_{t'=0}^t (T * E_{\tau_S})(t-t') dt'. \quad (10)$$

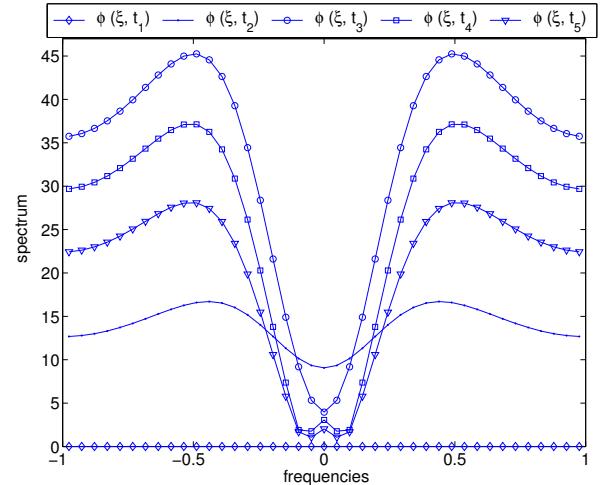


Fig. 2: The non-SPAM filter is a 2D spatially symmetric function. This figure shows a transversal cut of its spectrum for 5 different time samples of $R_c(t)$ and $R_s(t)$.

The above proposition is crucial for the reason that it enables the simplification and representation of the non-SPAM filter like a time-varying DoG. The DoG filters have

been extensively studied in the past [10, 21, 22]. Proposition 2.1 shows that the retinal-inspired filter can be modeled by a spatial DoG filter which is multiplied by the temporal filters $R_c(t)$ and $R_s(t)$ (Fig. 1), which act like weights and modify its spatial spectrum with respect to time (Fig. 2). The non-SPAM filter is shown in Fig. 2 where the parameters have been tuned according to neuroscientific results which approximates the retinal spectrum and the speed of the retinal processing: $\tau_C = 20.10^{-3}$, $\tau_s = 4.10^{-3}$, $\tau_G = 5.10^{-3}$, $w_S = 1$, $w_C = 0.75$, $\sigma_c = 0.5$, $\sigma_s = 1.5$. One can notice that, after a while, both the temporal filters $R_s(t)$ and $R_c(t)$ converge to the same value, which is established in the following proposition.

Proposition 2.2. *The filter $\phi(x, t)$ is a continuous and infinitely differential function such that $\phi(x, 0) = 0$ and*

$$\lim_{t \rightarrow +\infty} \phi(x, t) = \phi(x)$$

where $\phi(x)$ is a DoG filter.

In practice, $\phi(x, t)$ almost converges within a short time delay Δt . Hence, we assume that $\phi(x, \Delta t) \approx \phi(x)$ for all x . Hence, the non-SPAM filter only evolves during the time interval $[0, \Delta t]$. The time when this convergence almost occurs marks the end of the spatial evolution of the non-SPAM filter. In other words, the non-SPAM filter is capable to decompose the input image into different spatial subbands and extract information for a certain period of time Δt . After Δt all the necessary information has already been selected (Fig. 3).

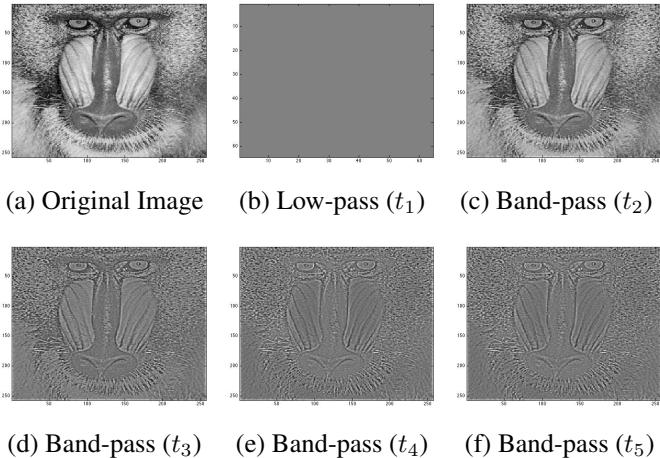


Fig. 3: Image decomposition by the non-SPAM filter.

3. NON-SPAM FRAME

The goal of this section is to prove that the non-SPAM filter is invertible such that we are able to reconstruct the input image. For this reason, we establish that the non-SPAM filter has a frame structure [22, 23]. For numerical purpose, we

need to discretize the non-SPAM filter. Let $x_1, \dots, x_n \in \mathbb{R}^2$ and $t_1, \dots, t_m \in \mathbb{R}^+$ be some sets of spatial and temporal sampling points. As a consequence, the continuous spatial convolution is approximated by the discrete convolution:

$$\begin{aligned} A(x_k, t_j) &= \phi(x_k, t_j) \otimes f(x_k) \\ &= \sum_{i=1}^n \phi(x_k - x_i, t_j) f(x_i), \end{aligned} \quad (11)$$

for all k and j . Let $\varphi_{k,j}$ be the row vector of \mathbb{R}^n defined by

$$\varphi_{k,j} = (\phi(x_k - x_1, t_j), \dots, \phi(x_k - x_n, t_j))$$

and Φ be the family of all these vectors:

$$\Phi = \left\{ \varphi_{k,j} \right\}_{1 \leq k \leq n, 1 \leq j \leq m}. \quad (12)$$

Let $f = (f(x_1), \dots, f(x_n))$ be the discretized image and $\|f\|$ its Euclidean norm. Let us denote $\tilde{\phi}_{t_j}(\xi)$ the discrete Fourier transform of the vector $(\phi(x_1, t_j), \dots, \phi(x_n, t_j))$.

Proposition 3.1. *The family of vectors Φ is a frame i.e. there exist two scalars $0 < \alpha \leq \beta < \infty$ such that:*

$$\alpha \|f\|^2 \leq \sum_{j=1}^m \sum_{k=1}^n |A(x_k, t_j)|^2 \leq \beta \|f\|^2, \quad (13)$$

where

$$\begin{aligned} \alpha &= \min_{\xi} \left\{ \frac{1}{n} \sum_{j=1}^m \left| \tilde{\phi}_{t_j}(\xi) \right|^2 \right\} > 0, \\ \beta &= \sum_{j=1}^m \sum_{k=1}^n \sum_{i=1}^n \phi^2(x_k - x_i, t_j). \end{aligned}$$

The proof of Proposition 3.1 is omitted due to the lack of place.

4. PROGRESSIVE RECONSTRUCTION

The progressive reconstruction consists in computing an estimate \hat{f}_{t_m} of the discretized image f at time t_m by using a limited amount of coefficients. In fact, at time t_m , all the coefficients of the non-SPAM frame has been computed but the reconstruction only exploits a small amount of them. We use the term progressive because the quality of the reconstruction increases as the amount of the coefficients in use increases.

When the decoder knows the total number of coefficients, he can reconstruct perfectly the input signal. Let us define $\tilde{A} = [\tilde{A}_{t_1}, \dots, \tilde{A}_{t_m}]$ as a vector of size nm and $\Phi = [\phi_1, \dots, \phi_m]$ a matrix of size $nm \times n$, where

$$\tilde{A}_{t_j} = \begin{bmatrix} \tilde{A}(x_1, t_j) \\ \vdots \\ \tilde{A}(x_n, t_j) \end{bmatrix} \quad \text{and} \quad \phi_j = \begin{bmatrix} \varphi_{1,j} \\ \vdots \\ \varphi_{n,j} \end{bmatrix}. \quad (14)$$

At time t_m , the estimate \hat{f}_{t_m} is given by

$$\tilde{f}_{t_m} = (\Phi^\top \Phi)^{-1} \Phi^\top \tilde{A}, \quad (15)$$

where M^{-1} denotes the inverse of a matrix M and M^\top denotes its transpose. The dual frame, which is necessary to have a perfect decoding at time t_m [22, 23], is $(\Phi^\top \Phi)^{-1} \Phi^\top$. Instead of computing the above matrix operator which can be time consuming and resource demanding, we can note that (15) is a solution of the following least squares problem:

$$\tilde{f}_{t_m} = \arg \min_{f \in \mathbb{R}^n} \left(\sum_{j=1}^m \|\phi_j \circledast f - \tilde{A}_{t_j}\|^2 \right). \quad (16)$$

This minimization problem can be easily solved by using a gradient descent algorithm.

The progressive reconstruction of the decoder is computed by selecting for each time bin the same percentages of spatial coefficients which are used to produce the reconstructed image. The decoder sorts in a descending order the total amount of coefficients for each time bin, and then it extracts the coefficients with the highest energy omitting the rest of the data. Obviously, the numerical results show that the reconstruction is better as the number of extracted coefficients is larger. This approach is based on the Rank-Order-Coding (ROC) model which is proposed in [24].

5. RESULTS

This section gives the numerical results of the progressive decoding introduced in the previous section. The parameters of the simulation are given in Fig. 2. The image is composed of 64×64 pixels. We define $\text{MSE}(f, \hat{f}_{t_m}) = \|f - \hat{f}_{t_m}\|^2/n$ the mean square error which measures the distortion between the original image f and the reconstructed image \hat{f}_{t_m} .

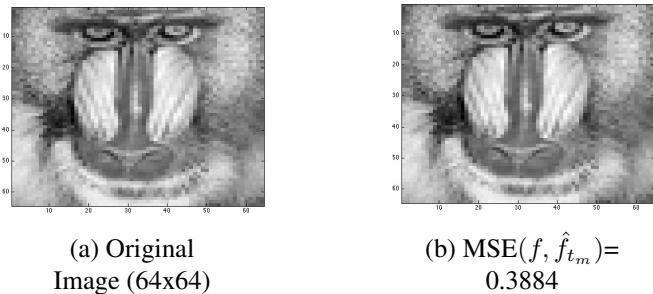


Fig. 4: Frame-based reconstruction.

Fig. 4 shows the optimal reconstruction when all the coefficients are used. The progressive reconstruction which is based on the ROC model is given in Fig. 5. It is obvious that while the number/percentage of coefficients increases, the quality of the reconstruction improves. However, even

for a very small percentage of the total amount of coefficients (i.e., 40%), the sharpness of the reconstructed image allows to recognize the basic structures of the input image.

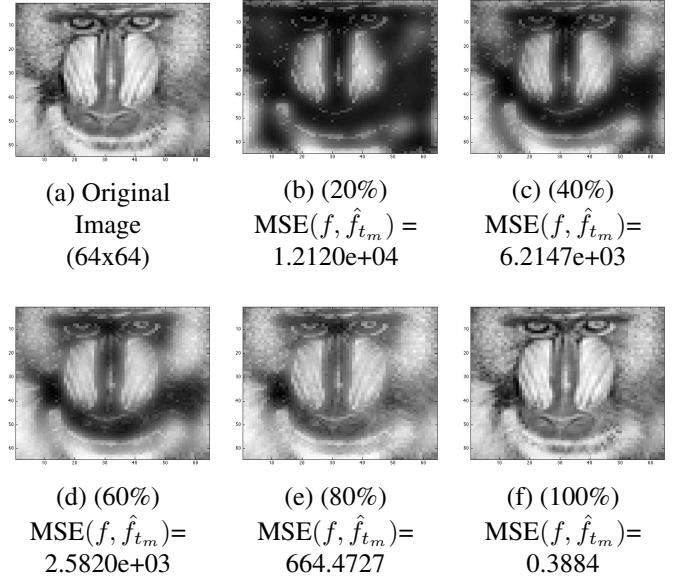


Fig. 5: Progressive reconstruction based on ROC model.

6. CONCLUSION

In this document, we have introduced a novel non-separable spatiotemporal filter (non-SPAM) based on a retinal model. This filter has a time-varying behavior. The progressive reconstruction exploits the rank-order-coding model to reconstruct the image by using a limited number of coefficients. For further study, we aim to extend this filtering result on a video stream.

7. REFERENCES

- [1] C. Poynton, *Digital Video and HDTV Algorithms and Interfaces*, Morgan Kaufmann Publishers, 2003.
- [2] A. Mulayoff D. Grois, D. Marpe and O. Hadar, “Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders,” *30th Picture Coding Symposium 2013 (PCS 2013)*, December 2013.
- [3] W. Gao, Y. Tian, T. Huang, S. Ma, and X. Zhang, “IEEE 1857 standard empowering smart video surveillance systems,” *IEEE Intelligent Systems*, 2013.
- [4] I. E. Richardson, *The H.264 Advanced Video Compression Standard*, 2011.
- [5] A. Wohrer and P. Kornprobst, “Virtual retina: A biological retina model and simulator, with contrast gain

control.,” *Journal of Computational Neuroscience*, vol. 26, no. 2, pp. 219–249, 2009.

[6] T. Gollisch and M. Meister, “Eye smarter than scientists believed: Neural computations in circuits of the retina,” *Neuron*, vol. 65, no. 2, pp. 150–164, 2010.

[7] R. Masland, “The fundamental plan of the retina.,” *Natural Neuroscience*, vol. 4, no. 9, pp. 877–886, 2001.

[8] K. Masmoudi, M. Antonini, and P. Kornprobst, “Streaming an image through the eye: The retina seen as a dithered scalable image coder,” *Signal processing Image Communication*, vol. 28, no. 8, pp. 856–869, 2013.

[9] P. Burt and E. Andelson, “The Laplacian pyramid as a compact image code,” *IEEE Trans. Commun*, vol. 31, no. 4, pp. 532–540, 1983.

[10] David J. Field, “What is the goal of sensory coding?”, *Neural Computation*, vol. 6, pp. 559–601, 1994.

[11] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 2nd edition, 1999.

[12] N. Ahmed, T. Natarajan, and K. R. Rao., “Discrete cosine transform,” *IEEE Transactions on Computers*, vol. 100, no. 1, pp. 90–93, 1974.

[13] I. Daubechies, “The wavelet transform, time-frequency localization and signal analysis.,” *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[14] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies., “Image coding using wavelet transform,” *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 205–220, 1992.

[15] M. Vetterli and J. Kovačević, *Wavelets and subband coding*, 1995.

[16] D. Gupta and S. Choubey, “Discrete wavelet transform for image processing,” *International Journal of Emerging Technology and Advanced Engineering*, vol. 4, no. 3, pp. 598–602, 2015.

[17] A.B. Watson, “The cortex transform: rapid computation of simulated neural images,” *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. , pages 311–327, 311–327, 1987.

[18] E. Andelson, E. Simoncelli, and R. Hingorani, “Orthogonal pyramid transforms for image coding,” *SPIE proceedings on Visual Communication and Image processing II*, vol. 845, pp. 50–58, 1988.

[19] J. P. Leduc, F. Mujica, R. Murenzi, and M. Smith, “Spatio-temporal wavelet transforms for motion tracking,” *ICASSP-97 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 3013 – 3016, 1997.

[20] J. G. Choi, S. Lee, and S. Kim, “Spatio-temporal video segmentation using a joint similarity measure,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 279 – 286, 1997.

[21] G. DeAngelis D. Cai and R. Freeman, “Spatiotemporal receptive field organization in the Lateral Geniculate Nucleus of cats and kittens,” *The American Physiological Society*, vol. 22, no. 3077, pp. 1045–1061, 1997.

[22] K. Masmoudi, M. Antonini, and P. Kornprobst, “Frames for exact inversion of the rank order coder,” *IEEE Transaction on Neural Networks*, vol. 23, no. 2, pp. 353–359, February 2012.

[23] J. Kovacevic and A. Chebina, “An introduction to frames,” *Signal Processing*, vol. 2, no. 1, pp. 1–94, 2008.

[24] R. Van Rullen and S. J. Thorpe, “Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex,” *Neural Computation*, vol. 13, pp. 1255–1283, 2001.